

# Molecular Cloning and Characterization of a New Member of the Gap Junction Gene Family, Connexin-31\*

(Received for publication, September 24, 1991)

Jan H. Hoh, Scott A. John, and Jean-Paul Revel‡

From the Division of Biology 156-29, California Institute of Technology, Pasadena, California 91125

A new member of the connexin gene family has been identified and designated rat connexin-31 (Cx31) based on its predicted molecular mass of 30,960 daltons. Cx31 is 270 amino acids long and is coded for by a single copy gene. It is expressed as a 1.7-kilobase mRNA that is detected in placenta, Harderian gland, skin, and eye. Cx31 is highly conserved and can be detected in species as distantly related to rat as *Xenopus laevis*. It exhibits extensive sequence similarity to the previously identified connexins, 58, 50, and 40% amino acid identity to Cx26, Cx32, and Cx43, respectively. When conservation of predicted phosphorylation sites is used to adjust the alignment of Cx31 to other connexins, a unique alignment of three predicted protein kinase C phosphorylation sites near the carboxyl terminus of Cx31 with three sites at the carboxyl terminus of Cx43 is revealed.

The gap junction is a structure composed of two closely apposed plasma membranes with a tightly packed array of cell to cell channels (Revel and Karnovsky, 1967). The physiology of the channels has been characterized in some detail in several experimental systems (Loewenstein, 1981; Spray and Bennett, 1985). In vertebrates, they have been shown to provide a low resistance electrical pathway between cells and to allow the passage of molecules <1000 Da with little or no selectivity (Flagg-Newton *et al.*, 1979). These channels are thought to have many important biological functions including the regulation of growth control (Mehta *et al.*, 1986), synchronization of cellular activity including synchronized contraction of myocardial cells (Barr *et al.*, 1965), regulation of embryonic development and differentiation (Pitts, 1978), and metabolic homeostasis (Sheridan *et al.*, 1979).

Gap junctions were identified first in the gold fish Mauthner cell by Robertson (Robertson, 1963). They have now been identified in almost every metazoan that has been examined, and they have also been described morphologically in a wide variety of tissues. This wide distribution and conservation of structure suggests that it is involved in a fundamental biological function shared by all multicellular animals.

Despite their wide distribution, gap junctions have been isolated only from a few organs in a few organisms. Hepatic gap junctions from mouse and rat were isolated, and two

proteins of  $M_r$  21,000 and  $M_r$  28,000 were identified as major structural components (Finbow *et al.*, 1980; Henderson *et al.*, 1979). Cardiac gap junctions isolated from rat, rabbit, and mouse have been shown to have a principal protein of  $M_r$  47,000 (Kensler and Goodenough, 1980; Manjunath *et al.*, 1982 and 1984). Gap junctions have also been isolated from rat and bovine lens, and a protein of  $M_r$  70,000 has been identified as its major component in the lens fiber cell (Kistler *et al.*, 1988). This protein has the same amino-terminal sequence as a protein with a predicted molecular mass of 46,000 Da (Beyer *et al.*, 1988). The relationship between the two proteins is not yet understood. Finally, gap junctions from the arthropods *Nephrops* and *Drosophila* have been isolated, and several putative protein components have been identified (Berdan and Gilula, 1988; Buultjens *et al.*, 1988; Ryerse, 1989).

The major structural proteins of gap junctions identified by isolation are now called connexins. They are members of a gene family that was first identified on the basis of protein sequence (Nicholson *et al.*, 1985; Nicholson *et al.*, 1981). Several cDNAs coding for connexins now have been isolated. In rat, the three proteins for which cDNAs have been described are designated Cx26,<sup>1</sup> Cx32, and Cx43 based on their predicted molecular mass (Beyer *et al.*, 1987; Paul, 1986; Zhang and Nicholson, 1989). These cDNAs correspond to the hepatic  $M_r$  21,000 and  $M_r$  28,000 proteins and the cardiac  $M_r$  47,000 protein, respectively. Homologues to these connexins from several other species have been isolated also (Gimlich *et al.*, 1990; Lash *et al.*, 1990; Musil *et al.*, 1990). Another member of the connexin gene family expressed in early *Xenopus* development, Cx38, has also been identified (Ebihara *et al.*, 1989). Through the use of the cDNAs as probes, the distribution of the various connexins has been described in several tissues and cell lines (Beyer *et al.*, 1987; Crow, *et al.*, 1990; Larson *et al.*, 1990; Musil *et al.*, 1990; Zhang and Nicholson, 1989). All connexins identified to date have unique distributions. Some tissues and organs have a single connexin while others have more than one, but no two are always found together. In rodent hepatocytes it appears that a single cell coexpresses both Cx26 and Cx32 (Nicholson *et al.*, 1987; Traub *et al.*, 1989).

Currently there is little known about the genes that code for the connexins. The gene coding for one connexin, rat Cx32, has been isolated and described (Miller *et al.*, 1988). It is a single copy gene with two exons, of which the second contains the entire coding region for the protein. The genes coding for Cx26 and Cx43 are also single copy (Musil *et al.*,

\* This work was supported in part by Grants HL37109 and BRSG-RR07003 from the National Institutes of Health and the Ruddock Fund. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) M59936.

‡ To whom correspondence should be addressed.

<sup>1</sup> The abbreviations used are: all connexins are abbreviated Cx followed by their predicted molecular mass. When appropriate, the species from which the connexin was identified is designated by the prefixes C, H, R, or X for chicken, human, rat, and *Xenopus*, respectively. Divergence times are given in million years (MYr) or billion years (BYr). Every 1000 bases is a kilobase (kb). DNA melting temperature is designated  $T_m$ .

1990; Zhang and Nicholson, 1989). The gene coding for Cx26 appears to have at least two exons, one of which contains the entire coding region as determined by Southern blotting (Zhang and Nicholson, 1989).

All connexins cloned to date were isolated first from cDNA libraries. The current study uses instead low stringency screening of a rat genomic library to isolate genes that code for connexin homologues to understand further the diversity, distribution, and phylogeny of this family of proteins. Using this alternate strategy we have identified a new member of the connexin gene family designated Cx31. Characterization of the gene reveals a 270-amino acid open reading frame with a high degree of sequence similarity to other connexins. This connexin shares many features with previously identified members of the family, including an alignment and conservation of three potential phosphorylation sites at the extreme carboxyl terminus of Cx43. We have also produced the first phylogenetic tree for the known connexins, which shows that Cx31 is closely related to Cx26 and Cx32 and distantly related to Cx43.

#### EXPERIMENTAL PROCEDURES

**Materials**—Restriction endonucleases and T7 RNA polymerase were obtained from Boehringer Mannheim. Modified T7 DNA polymerase and sequencing reagents were purchased from U. S. Biochemical Corp. Chemicals were from Sigma or Boehringer Mannheim. Sprague-Dawley rats were supplied by Simonsen Laboratories (Gilroy, CA).

**Isolation of Connexin Homologue Gene**—To identify connexin homologues we screened a Sprague-Dawley rat genomic *EcoRI* partial digest Charon 4A library using the Cx32 cDNA as a probe under low stringency conditions (Paul, 1986; Sargent *et al.*, 1979). Briefly, the Cx32 cDNA in pGEM-3 (Promega Biotec, Madison, WI) was linearized in the polylinker on the 5' side of the insert with *Bam*HI, and an antisense RNA probe with specific activity 2.5  $\mu$ Ci/ng was transcribed using T7 RNA polymerase and [ $\alpha$ -<sup>32</sup>P]UTP (Melton *et al.*, 1984; Tabor and Richardson, 1985). The RNA probe was used to screen approximately five genome equivalents ( $1 \times 10^6$  clones) of the genomic library plated at  $5 \times 10^4$  plaques per 150-mm plate and replicated in duplicate onto Hybond-N (Amersham Corp.) membranes. Prehybridization and hybridization were carried out in  $5 \times$  SSPE ( $1 \times$  SSPE = 150 mM NaCl, 10 mM Na<sub>2</sub>H(PO<sub>4</sub>), and 0.1 mM EDTA, pH 7.2),  $5 \times$  Denhardt's solution ( $50 \times$  Denhardt's = 1% w/v bovine serum albumin, 1% w/v polyvinylpyrrolidone, and 1% w/v Ficoll), 30% deionized formamide, 20  $\mu$ g/ml poly(A) RNA, 30  $\mu$ g/ml yeast tRNA, and 0.5–2 ng/ml RNA probe at 45 °C for 6 and 36 h, respectively. Membranes were treated at increasing stringency to a final wash of  $1 \times$  SSPE at 60 °C for 4 h and subjected to autoradiography. Positive clones were identified, and the procedure was repeated until clonal purity was achieved. The genomic clones isolated were characterized further by restriction site analysis and Southern blotting using the rat Cx26, Cx32, and Cx43 cDNAs as probes (Beyer *et al.*, 1987; Paul, 1986; Zhang and Nicholson, 1989). A 4.4-kb *EcoRI* fragment from one  $\lambda$  clone, RGJ21, which cross-hybridized with all three probes, was subcloned into pBluescript II KS(+) (Stratagene Cloning Systems, La Jolla, CA), characterized in more detail, and both strands of a 1.1-kb *EcoRI*-*SacII* fragment were sequenced by standard dideoxy sequencing (Sanger *et al.*, 1977; Tabor and Richardson, 1987).

**Genomic Southern Blots**—Southern blots of genomic DNA from rat (Sprague-Dawley), mouse (Balb/c), pig, and frog (*Xenopus laevis*) were performed under high and moderate stringency conditions. The rat DNA was isolated as described by Strauss (Strauss, 1988), the mouse and pig DNA were obtained commercially (Clontech, Palo Alto, CA), and the *Xenopus* DNA was provided by R. Wagner (California Institute of Technology). The rat DNA was digested separately with *Bgl*II, *Eco*RI, *Hind*III, *Kpn*I, *Nhe*I, *Sac*II or *Xba*I, separated on a 0.8% agarose gel, and capillary transferred onto Hybond-N as described by Maniatis *et al.* (1982). The blot was probed with the Cx31 *EcoRI*-*SacII* fragment random primer labeled to a specific activity of  $1 \times 10^9$  cpm/ $\mu$ g with [ $\alpha$ -<sup>32</sup>P]dATP (Feinberg and Vogelstein, 1984). High stringency prehybridization and hybridization were carried out in  $5 \times$  SSPE,  $5 \times$  Denhardt's solution, 50% deionized formamide, 1 mM sodium pyrophosphate, 1 mM ATP, 0.1% sodium dodecyl sulfate, 20  $\mu$ g/ml salmon sperm DNA, 30  $\mu$ g/ml yeast tRNA, and 2–5 ng/ml

probe DNA at 45 °C for 24 and 36 h, respectively. Blots were washed at increasing stringency to a final  $0.1 \times$  SSPE at 68 °C for 4 h. For the moderate stringency "zoo" blot all DNAs were digested with *Eco*RI or *Hind*III and treated as above, except that the hybridization conditions were  $5 \times$  SSPE,  $5 \times$  Denhardt's, 1 mM sodium pyrophosphate, 1 mM ATP, 0.1% sodium dodecyl sulfate, 20  $\mu$ g/ml salmon sperm DNA, 30  $\mu$ g/ml yeast tRNA, and 2–5 ng/ml probe DNA at 60 °C. The zoo blots were washed at increasing stringency to a final  $1 \times$  SSPE at 65 °C for 2–10 h.  $T_m$  values were estimated by the method of Meinkoth (Meinkoth and Wahl, 1984). All fragments were sized using the GEL<sup>2</sup> regression program.

**Northern Blots of Organ RNA**—To determine the expression pattern of RCx31 we isolated total RNA from several tissues and organs by the guanidinium thiocyanate method or the modified guanidinium thiocyanate method that includes a centrifugation through CsCl (MacDonald *et al.*, 1987). The sources of RNA were liver, heart, tail skin including the dermis, tail connective tissue (everything left after the skin was removed), Harderian gland, eye, placenta from a 19-day pregnant animal, epididymal fat pads, brain without the cerebellum, blood, stomach, femur including marrow, pancreas, spleen, ovary, uterus, thigh skeletal muscle, lung, duodenum, testis, and kidney, all from both male and female Sprague-Dawley rats. Northern blots were performed by electrophoresing 10–20 mg of each RNA in a formaldehyde gel. The separated RNA was capillary transferred onto Hybond-N and probed under conditions identical with the high stringency Southern blots except that the temperature was increased to 48–50 °C (Ausubel *et al.*, 1988).

**Analysis of RCx31 Sequence**—All computer analyses were carried out with either PC-Genie version 6.01 or 6.25 (PCG) running on an Epsom Equity II or the University of Wisconsin Genetics Computer Group package version 6.2 (GCG) (Devereux *et al.*, 1984) running on a VAX Station 3100 (Model m38) unless otherwise specified. All parameters are default unless otherwise indicated. Kyte and Doolittle hydropathy analysis was carried out on RCx31 using a 15-amino acid window in SOAP (PCG) (Kyte and Doolittle, 1982). The RCx31 sequence was scanned for known consensus sites for post-translational protein modification with PROSITE (PCG). Version 3.0 of PROSITE detects possible glycosylation, phosphorylation, sulfonation, amidation, fatty acylation, hydroxylation, carboxylation, phosphopantetheine attachment, and farnesyl group binding sites.

**Multiple Sequence Alignment**—Multiple alignments of the new rat Cx31 with protein sequences for rat Cx43, Cx32, and Cx26; chicken Cx43; *Xenopus* Cx43, Cx30, and Cx38; and human Cx32 were generated three sequences at a time using the program ALP3.<sup>3</sup> The ALP3 algorithm has been described by Murata *et al.* (1985). The triple alignments were compiled and adjusted by eye, taking into account the predicted phosphorylation sites, using the program LINEUP (GCG).

**Construction of Phylogenetic Tree**—To produce a phylogenetic tree and estimate divergence times, the connexins were divided into two homology domains; the first domain was from amino acid 2 to 99 in the Cx31, the second domain corresponded to positions 125–153, 155–165, and 168–211 in Cx31 and the aligned domains in the other connexins (see Fig. 6). Phylogenetic trees were generated using the program CLUSTAL (PCG) (Higgins and Sharp, 1988) for both domains I and II using rat Cx26, Cx31, Cx32, Cx43, and *Xenopus* Cx38. This program generates all possible pair-wise alignments using the algorithm of Wilbur and Lipman (1983). It then generates dendrograms using the unweighted pair group maximum averages method of Sneath and Sokal (1973). Divergence rates for Cx32 and Cx43 were estimated by linear regression analysis of plots of the percent corrected divergence for RCx32, HCx32, and XCx30; and RCx43, CCx43, and XCx43 versus divergence times using the method of Perler (Perler *et al.*, 1980). The times used are human/rodent 75 MYr, mammal/bird 275 MYr, and mammal/amphibian 350 MYr (Dayhoff, 1972; Doolittle *et al.*, 1989; Perler *et al.*, 1980).

#### RESULTS

**Isolation of a New Connexin Gene**—Thirty genomic clones that cross-hybridized to the RCx32 cDNA probe were isolated. Southern blotting and restriction site analysis of these clones

<sup>2</sup> GEL is a public domain program available from Intelligenetics, Mountain View, CA.

<sup>3</sup> ALP3 is a public domain program available from Intelligenetics, Mountain View, CA.



revealed that six corresponded to two different polymorphisms of the RCx32 gene.<sup>4</sup> A single clone designated RGJ21 that cross-hybridized with RCx26, RCx32, and RCx43 was identified and restriction-mapped (Fig. 1). The remaining clones are currently under further investigation, but three others that have been examined closely reveal no connexin homologues. The 4.4-kb fragment containing the homologous sequence was subcloned, and 982 bases between the *EcoRI*

and *SacII* sites were sequenced. Fig. 2 shows the nucleotide sequence and the translation of the nucleotide sequence that revealed a 270-amino acid open reading frame coding for a protein with a predicted molecular mass of 30,960 Da. The predicted protein exhibits a high degree of similarity to previously characterized connexins, 58, 50, and 40% amino acid identity, and 65, 58, and 51% nucleotide identity to RCx26, RCx32, and RCx43, respectively. Based on its homology to other connexins and predicted characteristics we designate this protein rat connexin-31 (RCx31).

**Genomic Southern Blots**—High stringency Southern blots of rat genomic DNA digested with seven different enzymes probed with Cx31 all revealed only a single band, including the 1.6-kb *KpnI* and 4.4-kb *EcoRI* fragments predicted from the genomic clone (Fig. 3). To control for cross-hybridization the same blot was hybridized with the RCx32 cDNA. It produced no bands identical with the Cx31 blot. This would suggest that RCx31 is a single copy gene. Southern blots of rat, mouse, pig, and frog genomic DNA were probed with Cx31 under conditions that allow sequences of greater than 75% nucleotide identity to be detected based on the estimated  $T_m$  (Fig. 4). Single *EcoRI* bands of 4.4, 4.1, 1.0, and 3.2 kb are seen in the rat, mouse, pig, and frog DNAs, respectively. Single *HindIII* bands of 3.0, 18, and 4.1 kb are also seen in the rat, mouse, and frog DNAs, respectively. In addition, a strong 3.3-kb *HindIII* band and three larger but weaker bands are seen in the pig DNA. This suggests that Cx31 is highly conserved and present in these species.

**Distribution of Cx31 mRNA**—A Northern blot of RNA

#### RGJ21

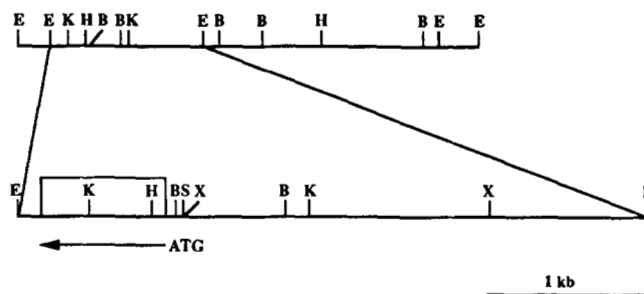


FIG. 1. Restriction map of the genomic clone RGJ21. The 4.4-kb *EcoRI* fragment that contains the gene coding for Cx31 is enlarged. The open reading frame is boxed, and the ATG start codon and coding direction of the protein are shown. The map was generated by single and double digests with the restriction endonucleases *EcoRI* (E), *BamHI* (B), *HindIII* (H), *KpnI* (K), *SacII* (S), and *XbaI* (X).

<sup>4</sup> Hoh, J. H., and Revel, J.-P. (1991) *Mamm. Genome*, in press.

FIG. 2. Nucleotide sequence and predicted protein sequence for Cx31. Predicted protein kinase C phosphorylation and casein kinase II sites as predicted by PROSITE are marked P and C, respectively.

```

GGTGAATCACTAGATAATGGGTTTTTCTCCATTGTTCCCTCAGGTGCACAGCCCCGAACCTGAGCAGGCACC
ATG GAT TGG AAG AAG CTT CAG GAC CTG TTG AGC GGT GTG AAC CAG TAC TCC ACC GCA TTC 60
MET Asp Trp Lys Lys Leu Gly Asn Leu Leu Ser Gly Val Asn Gln Tyr Ser Thr Ala Phe

GGG CGC ATC TGG CTG TCG GTA GTG TTC GTC TTC CGG GTG CTG GTG TAT GTG GTG GCT GCA 120
Gly Arg Ile Trp Leu Ser Val Val Phe Val Phe Arg Val Leu Val Tyr Val Val Ala Ala

GAG CGT GTG TGG GGC GAC GAG CAA AAA GAC TTT GAC TGT AAC ACC AGG CAG CCC GGT TGT 180
Glu Arg Val Trp Gly Asp Glu Gln Lys Asp Phe Asp Cys Asn Thr Arg Gln Pro Gly Cys

ACC AAC GTG TGC TAT GAC AAC TTC TTC CCC ATC TCC AAC ATC CGC CTC TGG GCC CTG CAG 240
Thr Asn Val Cys Tyr Asp Asn Phe Phe Pro Ile Ser Asn Ile Arg Leu Trp Ala Leu Gln

CTC ATC TTC GTC ACG TGT CCC TCT ATG CTG GTC ATC CTG CAC GTA GCC TAC CGC GAG GAG 300
Leu Ile Phe Val Thr Cys Pro Ser MET Leu Val Ile Leu His Val Ala Tyr Arg Glu Glu

CGG GAA CGG AAG CAT CGC CAG AAG CAC GGG GAG CAT TGC GCC AAA CTG TAC AGC CAC CCG 360
Arg Glu Arg Lys His Arg Gln Lys His Gly Glu His Cys Ala Lys Leu Tyr Ser His Pro

GGC AAG AAG CAC GGC GGC CTG TGG TGG ACC TAC CTG TTC AGT CTC ATC TTC AAG CTC ATC 420
Gly Lys Lys His Gly Gly Leu Trp Trp Thr Tyr Leu Phe Ser Leu Ile Phe Lys Leu Ile

ATT GAA TTG GTC TTC CTG TAT GTT CTA CAC ACG CTC TGG CAT GGC TTC ACC ATG CCG CGT 480
Ile Glu Leu Val Phe Leu Tyr Val Leu His Thr Leu Trp His Gly Phe Thr Met Pro Arg

CTG GTA CAG TGC GCC AGC GTG GTA CCT TGC CCC AAC ACC GTG GAT TGC TAC ATC GCT CGG 540
Leu Val Gln Cys Ala Ser Val Val Pro Cys Pro Asn Thr Val Asp Cys Tyr Ile Ala Arg

CCC ACG GAG AAG AAA GTC TTT ACC TAC TTC ATG GTA GGC GCG TCT GCC GTC TGC ATT ATT 600
Pro Thr Glu Lys Lys Val Phe Thr Tyr Phe Met Val Gly Ala Ser Ala Val Cys Ile Ile

CTC ACC ATC TGT GAG ATC TGC TAC CTC ATC TTC CAC AGG ATT ATG CGA GGC CTG AGC AAG 660
Leu Thr Ile Cys Glu Ile Cys Tyr Leu Ile Phe His Arg Ile Met Arg Gly Leu Ser Lys

GAC AAA TCG ACG AAG AGC ATC AGC TCC CCG AAG TCC TCC AGC CGG GCC TCC ACC TGT CGC 720
Asp Lys Ser Thr Lys Ser Ile Ser Ser Pro Lys Ser Ser Ser Arg Ala Ser Thr Cys Arg

TGT CAC CAC AAG CTG CTG GAG AGT GGT GAT CTG GAA GCA GTA CCA GCC GAT GAC AAG CTG 780
Cys His His Lys Leu Leu Glu Ser Gly Asp Leu Glu Ala Val Pro Ala Asp Asp Lys Leu

CAG GCT TCA GCG CCT AGC CTG ACC CCC ATT TAA CCACGCGCTGGAGAAGGGGTGAGGCTGGGAGGTG
Gln Ala Ser Ala Pro Ser Leu Thr Pro Ile ---

TGGAGGGGTCTGGGGTCTGAGTGCCCCCACTTTGAATTCCTGCAG

```

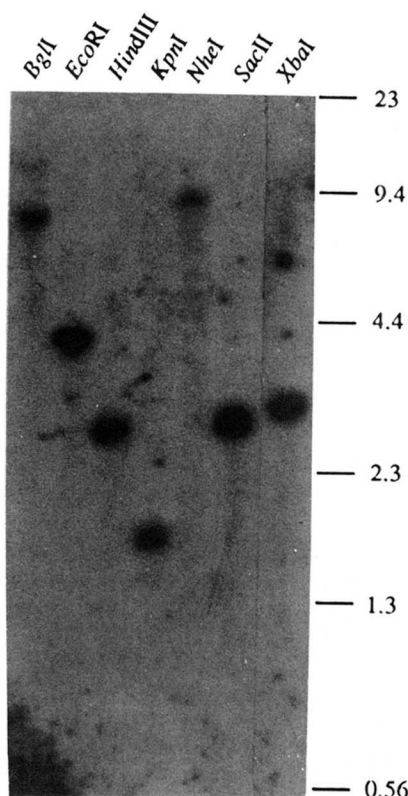


FIG. 3. Southern blot of rat genomic DNA probed with Cx31. The DNA was digested with various enzymes and probed at high stringency with a random primed Cx31 *EcoRI*-*SacII* fragment. Only a single band is seen in all lanes suggesting that Cx31 is a single copy gene. Hybridization of the same blot with a rat Cx32 probe revealed no bands identical with Cx31.

isolated from 20 different rat organs was probed with Cx31. It shows the presence of a 1.7-kb mRNA in placenta, Harderian gland, skin, and eye (Fig. 5). Some other organs show weak hybridization, but these cannot be described with confidence as positive signals. Control blots probed with RCx26, RCx32, and RCx43 show hybridization to 2.5-, 1.7-, and 3.0-kb mRNAs, respectively, in organs known to produce these transcripts, but no cross-hybridization with the RCx31 mRNA was detected. In addition to the organs known to express Cx26 and Cx43, we find a strong signal for both Cx26 and Cx43 in skin, and we detect Cx43 in bone, whole blood, and epididymal fat pads (data not shown). The control blots for Cx26 and Cx43 did not have RNA from placenta on them.

**Analysis of RCx31 Protein Sequence**—Hydropathy analysis of the RCx31 protein sequence reveals four highly hydrophobic amino acid segments similar to other connexins. In Cx32 and Cx43 these segments have been shown to be transmembrane spanning (Yancey *et al.*, 1989; Milks *et al.*, 1988; Zimmer *et al.*, 1987). Analysis of the protein for post-translational consensus modification sequences reveals potential phosphorylation and amidation sites. The protein kinase C consensus (S/T)-X-(RK) (Kishimoto *et al.*, 1985; Woodget *et al.*, 1986) is present at amino acid positions 182, 223, 229, 233, and 238 (Fig. 2). Position 182 is in the predicted second extracellular loop and would likely not be available to a cytoplasmic kinase. The other sites are all predicted to be exposed in the cytoplasm, in the carboxyl-terminal region. A single potential casein protein kinase II site is predicted at position 202, which is in the middle of the putative fourth transmembrane helix (Kuenzel and Mulligan, 1987). An amidation consensus site X-G-(RK)-(RK) is present at position 120 (Kreil, 1984).

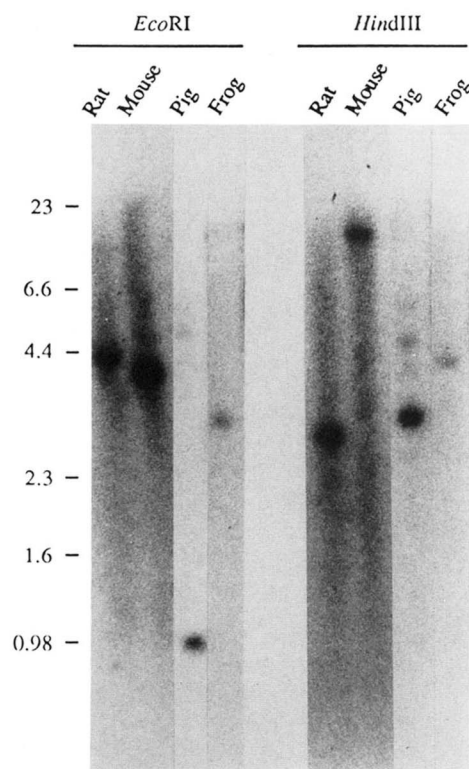


FIG. 4. Southern "zoo" blot of genomic DNA probed with Cx31. Genomic DNA from rat, mouse, pig, and frog was digested with *EcoRI* and *HindIII* and probed at moderate stringency with a random primed Cx31 *EcoRI*-*SacII* fragment. The figure is a composite from two different washes of the same blot. The rat and mouse lanes had excessive background after the first wash, and the pig and frog lanes were too weak to permit adequate reproduction after the second wash. All washes were done in 1 × SSPE at 65 °C.

**Multiple Alignment**—Multiple alignment of RCx31 with eight other connexins was carried out (Fig. 6; positions in the multiple alignment are referred to with the prefix MA). It includes all connexin homologues identified for which the full protein sequence is available, except for the bovine Cx43 recently published (Lash *et al.*, 1990). The alignment shows a perfect conservation of the three cysteines in the first extracellular loop, but RCx31 has a single amino acid inserted between the first two cysteines in the second extracellular loop. The cysteines have been shown to form at least one disulfide bond between the extracellular loops in Cx32<sup>5</sup> and Cx43.<sup>6</sup> RCx31 has a 22-amino acid deletion relative to the Cx43's cytoplasmic loop, leaving it with the smallest cytoplasmic loop of all the connexins. As is the case for Cx43s and XCx38, Cx31 has an arginine at MA162 in the third highly amphipathic putative transmembrane segment. The carboxyl-terminal regions of the different connexins are diverged highly. To improve the alignment of RCx31 over this region, we have taken into consideration the predicted post-translational modification sites. We find three predicted protein kinase C phosphorylation sites conserved when RCx31 and RCx43 are compared. These sites are at positions MA367, MA371, and MA376 in the multiple alignment and characterized by the sequence SSRAS. The third predicted phosphorylation site has a threonine substituted for a serine and is offset one amino acid from the corresponding Cx43 site that is at MA375. In addition the alignment shows two other sequence similarities among the other proteins that have not been

<sup>5</sup> B. Nicholson, personal communication.

<sup>6</sup> S. A. John, unpublished observations.



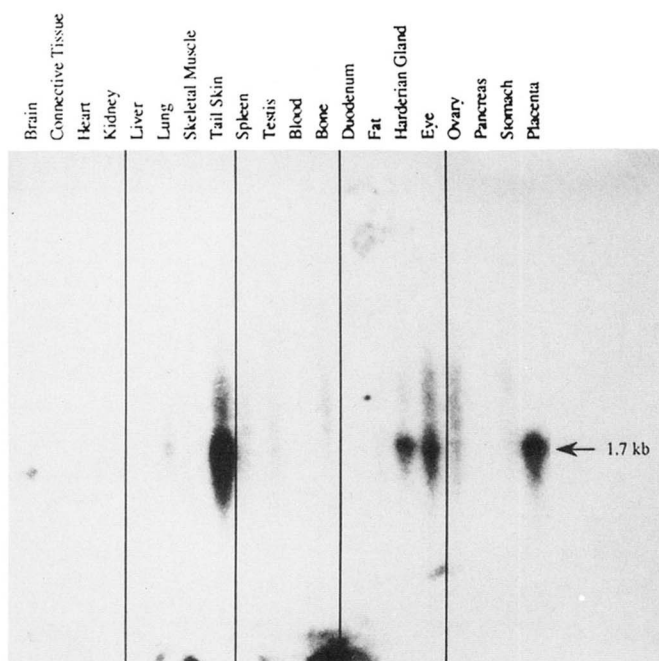


FIG. 5. Northern blot of total RNA isolated from 20 different rat organs probed with Cx31. RNAs were separated on a formaldehyde agarose gel and probed at high stringency with a random primed Cx31 *Eco*RI-*Sac*II fragment. The 1.7-kb mRNA band is detected in skin, Harderian gland, eye, and placenta. Control blots probed with Cx26, Cx32, and Cx43 did not cross-react with any Cx31 bands. The Cx32 mRNA comigrated with the Cx31 mRNA when the above blot was reprobed. However, Cx31 and Cx32 were not detected in any of the same tissues. Control blots did reveal the presence of Cx26 in skin, and of Cx43 in skin, bone, whole blood, and epididymal fat pads.

described previously. All connexins except RCx31 and RCx26 share a 11–13-amino acid stretch with four to six identities beginning at position MA319 in the multiple alignment characterized by a QNX<sub>6</sub>S sequence. Also, the carboxyl termini of Cx32 and Cx43 show a weak alignment across the SSRAS segment.

**Phylogenetic Tree of Connexins**—Using the program CLUSTAL, a tree of the five unique connexins identified to date, for which nucleotide sequences are available, was constructed (Fig. 7). The tree has two major branches, one with Cx38 and Cx43, and a second with Cx26, Cx31, and Cx32. Based on estimated divergence rates for Cx32 and Cx43, the two branches these molecules represent diverged 1.3–1.9 BYr ago.

#### DISCUSSION

All previously identified connexins were isolated as cDNAs from organ- or tissue-specific libraries. Because it has become clear that the connexins form a moderately large gene family and that the various connexins exhibit tissue specificity, we decided to use an alternate approach to isolate new connexin homologues by screening a rat genomic library at low stringency. This more general approach will identify new connexins irrespective of the tissue in which they are expressed. In this way we have identified so far one new connexin, rat Cx31, that we describe here. The gene for Cx32 was isolated six times from the five genome equivalents we screened, but only a single Cx31 gene was found. In addition, the low stringency screening did not identify genes for either Cx43 or Cx26. There are several possible explanations for these anomalies. The library we used was amplified, which is known to cause biases in representation. In addition, the library was generated from an *Eco*RI partial digest, which obviously will result in

some parts of the genome not being represented because the *Eco*RI fragments they generate are too large for the vector chosen.

The isolation of Cx31 now extends the number of connexins identified in rat to five, including Cx26, Cx32, Cx43, and Cx46. The nucleotide sequences for all of these, except Cx46, have been described. The number of genes isolated that code for connexins is now two, rat Cx31 and Cx32.

A significant issue with respect to Cx31 is whether the entire coding region for the protein for the gene is represented in the sequence from the single exon we describe. The answer to this cannot be known with certainty until a cDNA for Cx31 has been isolated and characterized. However, we believe that the coding region sequence is complete for the following reasons. The connexin genes coding for Cx26 and Cx32, for which there is sufficient structural detail, and that appear to be related closely to Cx31, both contain the entire coding region in a single exon. The mRNA size of 1.7 kb, although large enough to accommodate a larger coding region, is the same size as the Cx32 protein that is similar in size. Finally, there are no predicted exon/intron borders in Cx31 between the sequence coding for the fourth transmembrane helix and the carboxyl terminus (data not shown).

As in the case of the genes coding for rat Cx26, Cx32, and Cx43 (Miller *et al.*, 1988; Musil *et al.*, 1990; Zhang and Nicholson, 1989), the gene coding for Cx31 is a single copy gene. The fact that it shares no bands with Cx32 on a genomic Southern blot suggests that the Cx31 gene is not located near the Cx32 gene. Despite this, Cx31 and Cx32 may be part of a gene cluster with other connexins.<sup>7</sup> Further understanding of the Cx31 gene structure must await the isolation of a cDNA. By Southern blotting the Cx31 gene appears to be highly conserved, and a hybridization signal is seen in an organism as diverged from rat as the amphibian *X. laevis*. Of course it is possible that the signal seen in *Xenopus* with the Cx31 probe is a connexin isoform. However, all the connexins, Cx32 and Cx43, which have been isolated from multiple species are much more conserved between species than between isoforms. Further comparison of Cx32 and Cx43 sequences demonstrates that the amino-terminal region, corresponding to the first 99 amino acids, of these molecules is particularly well conserved and changes at a rate similar to the slowly evolving cytochrome *c* and glyceraldehyde-3-phosphate dehydrogenase (Dayhoff, 1972).

Cx31 has a unique distribution being found in skin, eye, Harderian gland, and placenta. It is found in the skin with Cx43 and Cx26. The skin is a complex tissue with many cell types including fibroblasts that have been shown to express Cx43 (Beyer *et al.*, 1989; Crow *et al.*, 1990). Intercellular communication in the epidermis has been studied in some detail (Kam *et al.*, 1986), but the exact localization of the different connexin molecules, their relationship to each other, and physiological roles have yet to be examined. Cx31 is found with Cx43 and MP70 (Cx46?) in the eye, where Cx43 is localized to the fibroblasts in the cornea and the cells of the lens epithelium (Beyer *et al.*, 1989) and MP70 is localized to the lens fiber cell (Kistler *et al.*, 1988).

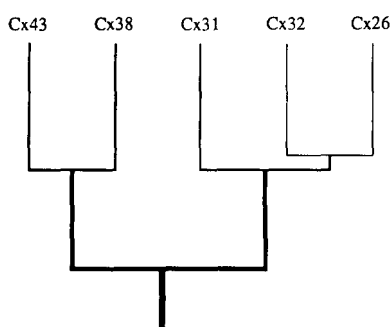
To date Cx31 is the only connexin found in the placenta. Gap junctions have been described in the mature rat placenta in all three layers of the labyrinth. The barrier between maternal and fetal blood supplies is at the boundary between layers II and III. Tracers applied through the maternal blood supply easily pass through layer I and accumulate at the border between II and III. These layers have been shown to contain gap junctions, and it has been proposed that these

<sup>7</sup> J. H. Hoh, unpublished observations.



**FIG. 6. Multiple alignment of rat connexins.** Rat connexin-31 (RCx31), -26 (RCx26), -32 (RCx32), and -43 (RCx43), *Xenopus* connexin-30 (XCx30), -38 (XCx38), and -43 (XCx43), chicken connexin-43 (CCx43), and human connexin-32 (HCx32). Consensus amino acids are shown in *upper case letters* and differences are shown in *bold lower case letters*. The criterion for a consensus is the identity of half, or four or more of the sequences at a given position. Alignments were carried out three at a time using the program ALP3, then compiled using LINEUP, and adjusted by eye. It shows the characteristic highly similar sequence for the first 230 amino acids except for a deletion of varying lengths with respect to all Cx43s at positions MA126–MA149. Further it reveals a conserved segment in several but not all connexins at MA320–MA333. When the alignment was adjusted further to take into account predicted phosphorylation sites, three conserved putative protein kinase C sites, marked with \*, were discovered in the RCx31 and RCx43 sequences. RCx32 also has a weak similarity across this region.

RCx31	M.DWklLqdl	LSGVNqYSTA	fGRlWLSVVF	vFRvLVyVVA	AErVWGDEQk	dFCNTzQPG	CtNVCyDnFF	PIShRLWsl	QLIFVtPam
RCx26	M.DWgtLqsl	LgGVNkHSTe	IGkIWLtLVF	IFRIMlLVVA	AkSVWGDEQa	dFCNTlQPG	CkNVCyDhYF	PISHRLWsl	QLlMVSTPAe
XCx30	M.nWaglyai	LSGVNzHSTe	IGRlWLSVVF	IFRIMVLVA	AESVWGDEks	AFcNTlQPG	CkNVCyDhFF	PISHRLWsl	QLlMVSTPAL
RCx32	M.nWtglytL	LSGVNzHSTe	IGRVLtSVlF	IFRIMVLVA	AESVWGDEks	AFcNTlQPG	CkNVCyDhFF	PISHRLWsl	QLlMVSTPAL
HCx32	M.nWtglytL	LSGVNzHSTe	IGRVLtSVlF	IFRIMVLVA	AESVWGDEks	AFcNTlQPG	CkNVCyDhFF	PISHRLWsl	QLlMVSTPAL
XCx38	MagWlLkll	LdVQVcHSTl	IGKVLtLVlF	IFRlFlLVA	GESVWtDEQS	dFCNTlQPG	CtNVCyDhYF	PlyHVRVWl	QLlFVSTPTL
XCx43	MGDwaLgkL	LdKvQaYSTA	gGKVWLSVlF	IFRlLLlgtA	VEsAWGDEQS	AFcNTlQPG	CkNVCyDhFF	PISHVRVWl	QLlFVSTPTL
CCx43	MGDwaLgkL	LdKvQaYSTA	gGKVWLSVlF	IFRlLLlgtA	VEsAWGDEhV	AFcNTlQPG	CkNVCyDhFF	PISHVRVWl	QLlFVSTPTL
RCx43	MGDwaLgkL	LdKvQaYSTA	gGKVWLSVlF	IFRlLLlgtA	VEsAWGDEQS	AFcNTlQPG	CkNVCyDhFF	PISHVRVWl	QLlFVSTPTL
RCx31	LVlLHVAYre	aeEzKhrqGh	gEhcaK.lys	hpgKKh	.....	GG	LWWTYlFsl	FKlIIElVFL	VylhtLwhGF
RCx26	LVAMHVAYrr	H.EKKrkKmk	gElKneKdI	eeIKtq	.....	.....	.....	.....	.....
XCx30	LVAMHVahq	HqEKKe..lr	lzhvkdqal	aeVKKh	.....	.....	.....	.....	.....
RCx32	LVAMHVahq	HlEKKm..lr	lghgqdlhl	eeVKKh	.....	.....	.....	.....	.....
HCx32	LVAMHVahq	HlEKKm..lr	lghgqdlhl	eeVKKh	.....	.....	.....	.....	.....
XCx38	tylgHmVYls	kKEEKerqKE	nEaIlvane	aqtevy	.....	.....	.....	.....	.....
XCx43	LYLAHVfYlm	rKEEKlnrKE	eeElkmvqng	gnvdmHLKQI	EIKKFKYGlE	EHGKVKmRG	LlTYIISl	FKSVFEVFL	LIQWYIY.GF
CCx43	LYLAHVfYlm	rKEEKlnrKE	eeElkmvqng	gnvdmHLKQI	EIKKFKYGlE	EHGKVKmRG	LlTYIISl	FKSVFEVFL	LIQWYIY.GF
RCx43	LYLAHVfYlm	rKEEKlnrKE	eeElkmvqng	gnvdmHLKQI	EIKKFKYGlE	EHGKVKmRG	LlTYIISl	FKSVFEVFL	LIQWYIY.GF
RCx31	tMPrLVGCaS	vVPCNTVDC	yIARPTeKKV	FTyFMvGaSa	VCIltIcEi	cYLIF.....	.....	HrimrglSkd	Kst.....
RCx26	fMqRLVKCa	awPCNTVDC	FISRPTEKTV	FTVFMlVSG	IClILNlTEL	cYLlFyCag	ksKRpV	.....	.....
XCx30	SMIRLLKCD	ayPCNTVDC	FVSRTEKTI	FTVFMlVSG	VCIvLnVaeV	FfIlaqaCtR	RaxR	.....	.....
RCx32	amVRLVKCa	afPCNTVDC	FVSRTEKTV	FTVFMlVSG	IClILNlTEL	vYLIraCaR	RaQRrSnPps	rkqSGfghrl	apeyk.....
HCx32	amVRLVKCD	vYPCNTVDC	FVSRTEKTV	FTVFMlVSG	IClILNlTEL	vYLIraCaR	RaQRrSnPps	rkqSGfghrl	apeyk.....
XCx38	vMSPifvCe	RlPCkHkVc	FVSRTEKTI	FtFmLVVSl	ISlILNlEL	ihlFkFqgh	gike	.....	gatcpPkgip
XCx43	SLsaiytCk	RdPCPHqVDC	FISRPTEKTI	FtFmLVVSl	VSLALNIIEl	FyvtKfakd	gikgkDPF	.....	aatndavag
CCx43	SLsaiytCk	RdPCPHqVDC	FISRPTEKTI	FtFmLVVSl	VSLALNIIEl	FyvtKfakd	RvkgktDPY	.....	shSGtmSpa
RCx43	SLsavytCk	RdPCPHqVDC	FISRPTEKTI	FtFmLVVSl	VSLALNIIEl	FyvtKfakd	RvkgRSDPY	HattOpIsPa	KdGSPKYAY
RCx31	.....	.....	.....	.....	.....	.....	.....	.....	360
RCx26	.....	.....	.....	.....	.....	.....	.....	.....	.....
XCx30	.....	.....	.....	.....	.....	.....	.....	.....	.....
RCx32	.....	.....	.....	.....	.....	.....	.....	.....	.....
HCx32	.....	.....	.....	.....	.....	.....	.....	.....	.....
XCx38	FNG.....A	gnrMppqet	npssnqddl	lpaYNKmgag	hNwssiqmE	qvnGlvkpkc	qcd.....	.....	ISVvvgagp
XCx43	FNGCSPSTA	..PMSPPGYK	LVTGERNPS	CRNYNKQASE	QNWANYSAEQ	NrMGqagStl	SNehAQPFDF	.....	DEHQNTKKa
CCx43	YNGCSPSTAP	LSPMSPPGYK	LVTGERNPS	CRNYNKQASE	QNWANYSAEQ	NrMGqagStl	SNShAQPFDF	.....	ASGHELOPLt
RCx43	FNGCSPSTAP	LSPMSPPGYK	LVTGERNPS	CRNYNKQASE	QNWANYSAEQ	NrMGqagStl	SNShAQPFDF	pDdnQNAKkv	ASGHELOPLa
RCx31	..ksIsapks	SSzASterch	hkLlESGDLE	AVPADDKLQA	SAPSLTPI	.....	.....	.....	.....
RCx26	.....	.....	.....	.....	.....	.....	.....	.....	.....
XCx30	..lIkRSag	.....	gqkgdnc	ats	.....	.....	.....	.....	.....
RCx32	dIlrRSpgtg	agLAeksdRc	sac	.....	.....	.....	.....	.....	.....
HCx32	dIlrRSpgtg	agLAeksdRc	sac	.....	.....	.....	.....	.....	.....
XCx38	IInmdavkr	nhqtSSkqry	v	.....	.....	.....	.....	.....	.....
XCx43	Apghemqpl	SSHASSrPp	ddLEI	.....	.....	.....	.....	.....	.....
CCx43	IvdqRppsa	SSzASSrPp	ddLEI	.....	.....	.....	.....	.....	.....
RCx43	IvdqRppsa	SSzASSrPp	ddLEI	.....	.....	.....	.....	.....	.....



**FIG. 7. Phylogenetic tree of connexins.** The tree is based on the CLUSTAL analysis of domains I and II and of the protein sequences for rat Cx26, Cx31, Cx32, and Cx43, and *Xenopus* Cx38. The estimated time for the divergence of the two major branches is 1.3–1.9 BYr ago.

gap junctions are responsible for the exchange of fluids, gases, and small metabolites between the mother and fetus in rat (Metz *et al.*, 1978).

The Harderian gland is a secretory organ located on the posterior aspect of the eye in animals with a third eye lid. The biological function of the gland is not well understood. In mammals it has been shown to consist of two major cell types, secretory cells and myoepithelial cells (Woodhouse and Rhodin, 1963). In the hamster, both these cell types are innervated, and it has been suggested that the secretory functioning of the gland is regulated neuronally (Bucana and Nadakavukaren, 1972). A neuronal signal that activates secretion might be transmitted electrically or biochemically between cells through gap junctions. Further study of Cx31,

which is the only connexin yet identified in the Harderian gland, may shed light on this process.

Analysis of the Cx31 protein sequence reveals several potential protein modification sites. The single potential amidation site is probably not utilized, because amidation usually is associated with processed proteins such as neuropeptides (Kreil, 1984). There are five predicted protein kinase C sites. The first, at position 182, is in the predicted second extracellular domain and probably not available to a kinase. The same is true for the single casein protein kinase II site at position 202, which is in the middle of the predicted fourth transmembrane helix. The remaining four protein kinase C sites are in the carboxyl-terminal region. Whether these sites are utilized must await biochemical analysis of the protein. It is interesting to note that protein kinase C and cAMP-dependent protein kinase have been implicated in regulation of junctional communication in a number of systems (Murray and Gainer, 1989). Cx32 has been shown to be phosphorylated by cAMP-dependent protein kinase *in vivo* and *in vitro* (Saez *et al.*, 1986; Takeda *et al.*, 1987). In addition, Cx43 in fibroblasts is phosphorylated (Crow *et al.*, 1990).

The multiple alignments (Fig. 6) show the general characteristics of the gene family that have been described in previous comparisons (Beyer *et al.*, 1990; Gimlich *et al.*, 1990). The molecules have a 200–220 amino acid core which is highly conserved, except for the intracellular loop. The loop is of variable length but highly conserved between homologous connexins from different species. The high degree of sequence similarity through the first 200–220 amino acids is particularly interesting considering the high degree of morphological conservation of the gap junction observed. Two notable features

in this region are the conserved cysteines and the putative amphipathic helix. The two sets of cysteines have the pattern CX<sub>5</sub>CX<sub>3</sub>C in the first extracellular loop and CX<sub>3</sub>CX<sub>5</sub>C in the second extracellular loop. The only exception is in Cx31 where the second set of cysteines has a single amino acid inserted which results in the pattern CX<sub>4</sub>CX<sub>5</sub>C. The putative amphipathic helix, which has been proposed to line the pore of the channel, begins at position MA154 and has the conserved motif TX<sub>3</sub>SX<sub>3</sub>K/RX<sub>3</sub>E. The carboxyl termini are highly diverged and vary in length from approximately 10 to 150 amino acids and may provide regulatory specificity to the various connexin isoforms. Several aspects of the alignment we show here are different from ones published previously. We find a unique alignment in the carboxyl-terminal region characterized by the sequence QNX<sub>6</sub>S or QNX<sub>9</sub>S, starting at position MA320, conserved in all connexins except RCx31 and RCx26. The significance of this sequence is not known, and it does not match any known structural or functional motif as determined by PROSITE. Its conservation in many connexins would suggest that it may be involved in a shared structure or function. Further study of RCx31 and its physiology may reveal more about the sequence, since it is not present in that molecule. The approach of using predicted phosphorylation sites to adjust the alignment of the sequences revealed a segment characterized by the sequence SSRAS, starting position at MA371, that is shared between RCx31 and RCx43. There is also a weak alignment with RCx32, which contains a predicted protein kinase C site at position MA377. The SSRAS segment has several putative protein kinase C phosphorylation sites in RCx31 or RCx43, but no actual positions of protein kinase C phosphorylation in any of these molecules has been reported. RCx32 has been shown to be phosphorylated *in vitro* by protein kinase C (Takeda *et al.*, 1989).

The phylogenetic tree of connexins generated from analysis of the divergence rates, and the CLUSTAL program, predicts a branch between Cx43/Cx38 and Cx26/Cx31/Cx32 at 1.3–1.9 BYr. That estimated divergence time assumes that the amino acid replacement method of Perler is an accurate clock (Perler *et al.*, 1980). It is relevant to note that the prokaryote/eukaryote divergence is estimated at 1.8 BYr, the plant/animal at 1.0 BYr, and the vertebrate/invertebrate at 0.6 BYr (Doolittle *et al.*, 1989). It is also interesting to note that the phylogenetic tree of connexins is consistent with the nomenclature proposed for the gap junction protein family by Gimlich *et al.* (1990). That nomenclature is based on unspecified sequence similarities between connexins and uses greek letters to identify homologous proteins. Eventually it will be useful to have a naming system independent of calculated molecular weight, because a name based on the size of the molecule will become useless once the repertoire of connexin isoforms and connexins from different species becomes too large. We suggest that phylogenetic relatedness would be a suitable criterion for naming connexins, but such a system must await further definition and analysis of the gene family. Until that time the current system of using the prefix Cx followed by the molecular mass is a practical approach.

**Acknowledgments**—We thank Dr. D. Paul for the Cx32 cDNA clone, Dr. E. Beyer for the Cx43 cDNA, Dr. B. Nicholson for the Cx26 cDNA, Dr. J. Bonner for the rat genomic library, and R. Wagner for the *Xenopus* genomic DNA. We are also grateful to Dr. B. Yancey and Dr. K. Puranam for comments and suggestions on the manuscript.

#### REFERENCES

- Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1988) *Current Protocols in Molecular Biology*, Greene and Associates and Wiley Interscience, New York
- Barr, L., Dewey, F. F., and Berger, W. (1965) *J. Gen. Physiol.* **48**, 797–823
- Berdan, R. C., and Gilula, N. B. (1988) *Cell Tissue Res.* **251**, 257–274
- Beyer, E. C., Paul, D. L., and Goodenough, D. A. (1987) *J. Cell Biol.* **105**, 2621–2629
- Beyer, E. C., Goodenough, D. A., and Paul, D. L. (1988) in *Gap Junctions* (Hertzberg, E. L., and Johnson, R. G., eds) Vol. 7, pp. 167–176, Alan R. Liss, Inc., New York
- Beyer, E. C., Kistler, J., Paul, D. L., and Goodenough, D. A. (1989) *J. Cell Biol.* **108**, 595–605
- Beyer, E. C., Paul, D. L., and Goodenough, D. A. (1990) *J. Membr. Biol.* **116**, 187–194
- Bucana, C. D., and Nadakavukaren, M. J. (1972) *Science* **175**, 205–206
- Bultjens, T. E. J., Finbow, M. E., Lane, N. J., and Pitts, J. D. (1988) *Cell Tissue Res.* **251**, 571–580
- Crow, D. S., Beyer, E. C., Paul, D. L., Kobe, S. S., and Lau, A. F. (1990) *Mol. Cell. Biol.* **10**, 1754–1763
- Dayhoff, M. O. (1972) *Atlas of Protein Sequence and Structure* 1972, The National Biomedical Research Foundation, Silver Spring, MD
- Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–394
- Doolittle, R. F., Anderson, K. L., and Feng, D.-F. (1989) in *The Hierarchy of Life* (Fernholm, B., Bremer, K., and Joernvall, H., eds) pp. 73–85, Excerpta Medica, New York
- Ebihara, L., Beyer, E. C., Swenson, K. I., Paul, D. L., and Goodenough, D. A. (1989) *Science* **243**, 1194–1195
- Feinberg, A. P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266–267
- Finbow, M. E., Yancey, S. B., Johnson, R. G., and Revel, J. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77**, 970–974
- Flagg-Newton, J., Simpson, I., and Loewenstein, W. R. (1979) *Science* **205**, 404–407
- Gimlich, R. L., Kumar, N., and Gilula, N. B. (1990) *J. Cell Biol.* **110**, 597–605
- Henderson, D., Eibl, H., and Weber, K. (1979) *J. Mol. Biol.* **132**, 193–218
- Higgins, D. G., and Sharp, P. M. (1988) *Gene (Amst.)* **73**, 237–244
- Kam, E., Melville, L., and Pitts, J. D. (1986) *J. Invest. Dermatol.* **87**, 748–756
- Kensler, R. W., and Goodenough, D. A. (1980) *J. Cell Biol.* **86**, 755–764
- Kishimoto, A., Nishiyama, K., Nakanishi, H., Uratsugi, Y., Nomura, H., Takeyama, Y., and Nishizuka, Y. (1985) *J. Biol. Chem.* **260**, 12492–12499
- Kistler, J., Christie, D., and Bullivant, S. (1988) *Nature* **331**, 721–723
- Kreil, G. (1984) *Methods Enzymol.* **106**, 218–223
- Kuenzel, E. A., Mulligan, J. A., Sommercorn, J., and Krebs, E. G. (1987) *J. Biol. Chem.* **262**, 9136–9140
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Larson, D. M., Haudenschield, C. C., and Beyer, E. C. (1990) *Circ. Res.* **66**, 1074–1080
- Lash, J. A., Critser, E. S., and Pressler, M. L. (1990) *J. Biol. Chem.* **265**, 13113–13117
- Loewenstein, W. R. (1981) *Physiol. Rev.* **61**, 829–913
- MacDonald, R. J., Swift, G. H., Przybyla, A. E., and Chirgwin, J. M. (1987) *Methods Enzymol.* **152**, 219–227
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Manjunath, C. K., Goings, G. E., and Page, E. (1982) *Biochem. J.* **205**, 189–194
- Manjunath, C. K., Goings, G. E., and Page, E. (1984) *Am. J. Physiol.* **246**, H865–H875
- Mehta, P., Bertram, J., and Loewenstein, W. R. (1986) *Cell* **44**, 187–196
- Meinkoth, J., and Wahl, G. (1984) *Anal. Biochem.* **138**, 267–284
- Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984) *Nucleic Acids Res.* **12**, 7035–7056
- Metz, J., Aoki, A., and Forsmann, W. G. (1978) *Cell Tissue Res.* **192**, 391–407
- Milks, L., Kumar, N. M., Houghton, R., Unwin, N., and Gilula, N. B. (1988) *EMBO J.* **7**, 2967–2975
- Miller, T., Dahl, G., and Werner, R. (1988) *Biosci. Rep.* **8**, 455–464

- Murata, M., Richardson, J. S., and Sussman, J. L. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 3073–3077
- Murray, A. and Gainer, H. S. C. (1989) in *Cell Interactions and Gap Junctions* (Sperelakis, N., and Cole, W. C., eds) Vol. 1, pp. 97–107, CRC Press, Boca Raton, FL
- Musil, L. S., Beyer, E. C., and Goodenough, D. A. (1990) *J. Membr. Biol.* **116**, 163–175
- Nicholson, B. J., Hunkapiller, M. W., Grim, L. B., Hood, L. E., and Revel, J.-P. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7594–7598
- Nicholson, B. J., Gros, D. B., Kent, S. B. H., Hood, L. E., and Revel, J.-P. (1985) *J. Biol. Chem.* **260**, 6514–6517
- Nicholson, B., Dermietzel, R., Teplow, D., Traub, O., Willecke, K., and Revel, J.-P. (1987) *Nature* **329**, 732–734
- Paul, D. L. (1986) *J. Cell Biol.* **103**, 123–134
- Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) *Cell* **20**, 555–566
- Pitts, J. D. (1978) in *Intercellular Junctions and Synapses, Receptors and Recognition, Series B* (Feldman, J., Gilula, N. B., and Pitts, J. D., eds) Vol. 2, pp. 63–79, Chapman and Hall, London
- Revel, J.-P., and Karnovsky, M. J. (1967) *J. Cell Biol.* **33**, C7–C12
- Robertson, J. D. (1963) *J. Cell Biol.* **19**, 201–221
- Ryerse, J. S. (1989) *Cell Tissue Res.* **256**, 7–16
- Saez, J., Spray, D. C., Nairn, A. C., Hertzberg, E. L., Greengard, P., and Bennett, M. V. L. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2473
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
- Sargent, T. D., Wu, J.-R., Sala-Trepat, J. M., Wallace, R. B., Reyes, A. A., and Bonner, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3256–3260
- Sheridan, J. D., Finbow, M. E., and Pitts, J. D. (1979) *Exp. Cell Res.* **123**, 111–117
- Sneath, P. H. A., and Sokal, R. R. (1973) *Numerical Taxonomy*, W. H. Freeman & Co., New York
- Spray, D. C., and Bennett, M. V. L. (1985) *Annu. Rev. Physiol.* **47**, 281–303
- Strauss, W. M. (1988) in *Current Protocols in Molecular Biology* (Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., and Stuhl, K., eds.) Vol. 1, pp. 2.2.1–2.2.3, Greene and Associates and Wiley Interscience, New York
- Tabor, S., and Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 1074–1078
- Tabor, S., and Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4767–4771
- Takeda, A., Hashimoto, E., Yamamura, H., and Shimazu, T. (1987) *FEBS Lett.* **210**, 169–172
- Takeda, A., Saheki, S., Shimazu, T., and Takeuchi, N. (1989) *J. Biochem.* **106**, 723–727
- Traub, O., Look, J., Dermietzel, R., Brummer, F., Hulser, D., and Willecke, K. (1989) *J. Cell Biol.* **108**, 1039–1051
- Wilbur, W. J., and Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 726–730
- Woodget, J. R., Gould, K. L., and Hunter, T. (1986) *Eur. J. Biochem.* **161**, 177–184
- Woodhouse, M. A., and Rhodin, J. A. G. (1963) *J. Ultrastruct. Res.* **9**, 76
- Yancey, S. B., John, S. A., Lal, R., Austin, B. J., and Revel, J.-P. (1989) *J. Cell Biol.* **108**, 2241–2254
- Zhang, J.-T., and Nicholson, B. J. (1989) *J. Cell Biol.* **109**, 3391–3401
- Zimmer, D., Green, C. R., Evans, W. H., and Gilula, N. B. (1987) *J. Biol. Chem.* **262**, 7751–7763